OCR
Oxford Cambridge and RSA

An OCR endorsed teaching and learning tool

OCR A Level Computer Science H446 – Paper 1

6

Problem recognition

Unit 10 Computational thinking

PG ONLINE

Objectives

- Learn about and apply the following to solve problems
  - Visualisation
  - Backtracking
  - Data mining
  - Heuristics
  - Performance modelling
  - Pipelining

Writing algorithms

- One of the oldest known algorithms was written by Euclid over 2000 years ago
- It’s designed to find the greatest common divisor of two numbers (the largest number which divides both of them)

The GCD algorithm

- Suppose x and y represent two whole numbers
- Euclid's algorithm goes like this:

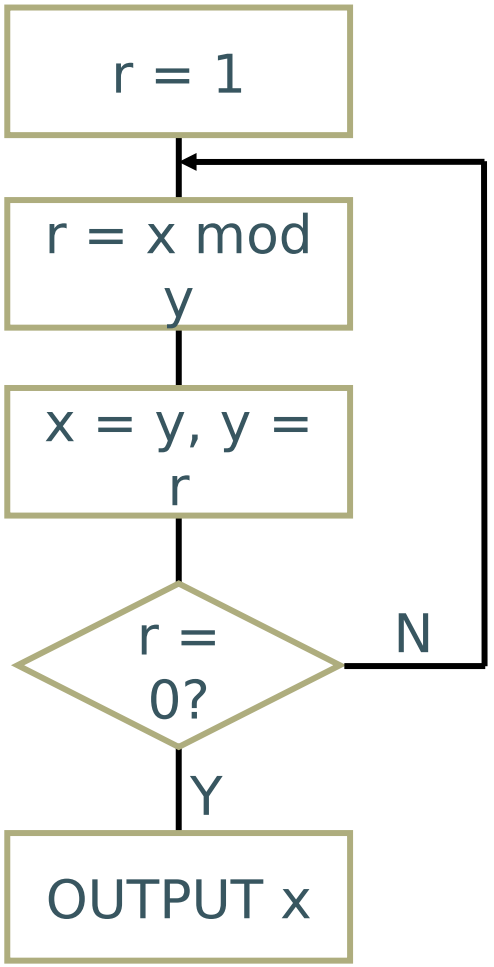
```
r = 1
WHILE r <> 0
    r = x mod y
    x = y
    y = r
ENDWHILE
OUTPUT y
```

- A flowchart is a visualisation of an algorithm
  - Draw a flowchart of this algorithm

The GCD algorithm

```
r = 1
WHILE r <> 0
    r = x mod y
    x = y
    y = r
ENDWHILE
OUTPUT x
```

- Test the algorithm by tracing the values of r, x and y starting with x = 420, y = 66



The GCD algorithm

```
r = 1
WHILE r <> 0
    r = x mod y
    x = y
    y = r
ENDWHILE
OUTPUT x
```

- Test the algorithm by tracing the values of r, x and y starting with x = 420, y = 66
  - Output is 6

r	x	y
1	420	66
24	66	24
18	24	18
6	18	6
0	6	0

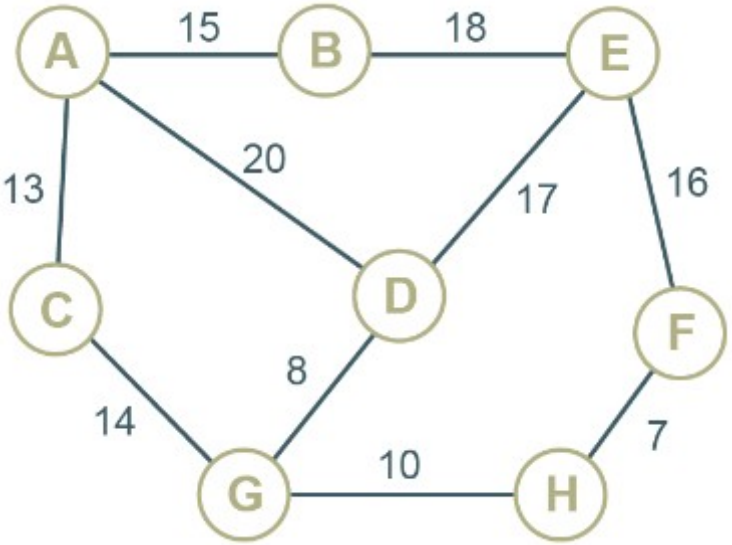
Visualising an algorithm

- An image is much easier for a human to understand than a table of numbers
- What does the graph represented below look like?

	A	B	C	D	E	F	G	H
A		15	13	20				
B	15				18			
C	13						14	
D	20				17		8	
E		18		17		16		
F					16			7
G			14	8				10
							1	

The graph

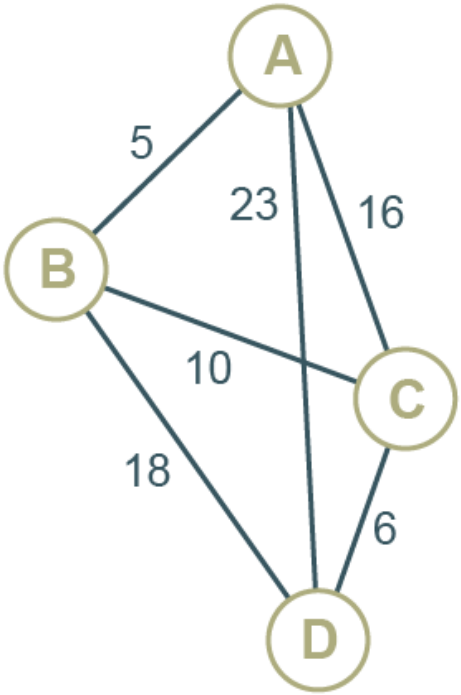
- Here it is visualised as a set of nodes and edges



- The weights on the edges could be distances or costs

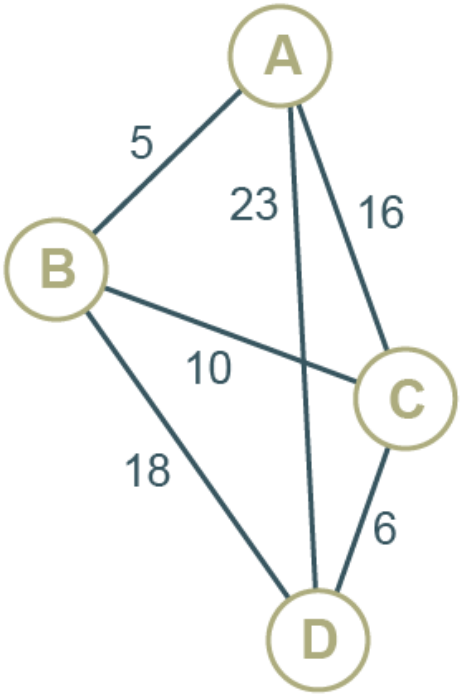
Shortest path algorithm

- Suppose we wanted to find the shortest path from city A to city D
- One method would be to perform an exhaustive search of all possible routes
- With 4 cities, how many possible routes are there?
  - What is the shortest route?



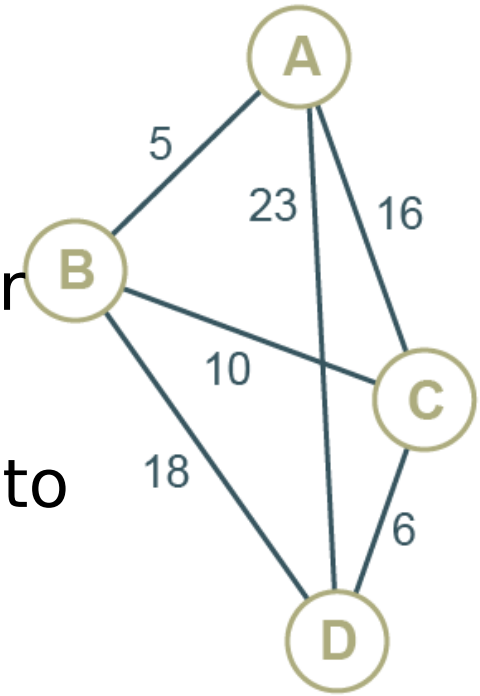
Exhaustive search

- With 4 cities, there are 3 x 2 = 6 possible routes
- With 5 cities, there are 4 x 3 x 2 routes
- The shortest route from A to D is A → B → C → D, a distance of 21
- With only 4 cities, an exhaustive search is possible
  - But what if there are 8 cities?



Exhaustive search

- The formula for the number of possible routes for 8 cities is: 7! = 7 x 6 x 5 x 4 x 3 x 2 x 1 = 211,680
- The number of possible routes for 10 cities is 10! = 3,628,800
- Clearly a better algorithm needs to be found
  - We'll use an algorithm to find the shortest distance from A to D

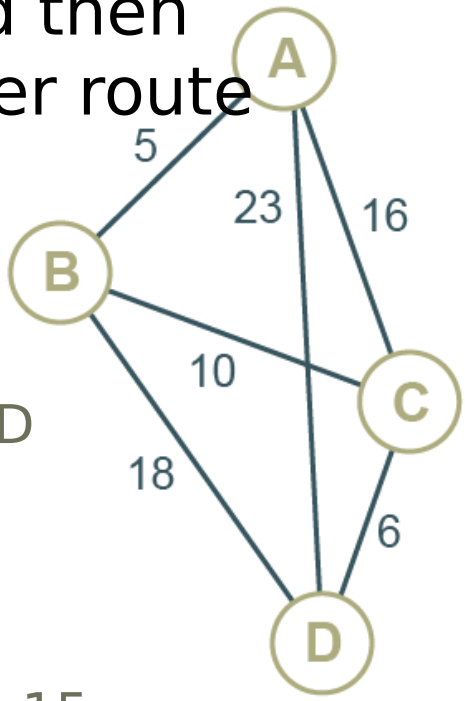


Backtracking

- One technique is backtracking, whereby you go some way along one route and then backtrack to see if there is a better route
  - Starting at A, visit B, C and D to see which is the shortest distance
  - B is shortest, so backtrack to A, travel to B and from there visit C and D

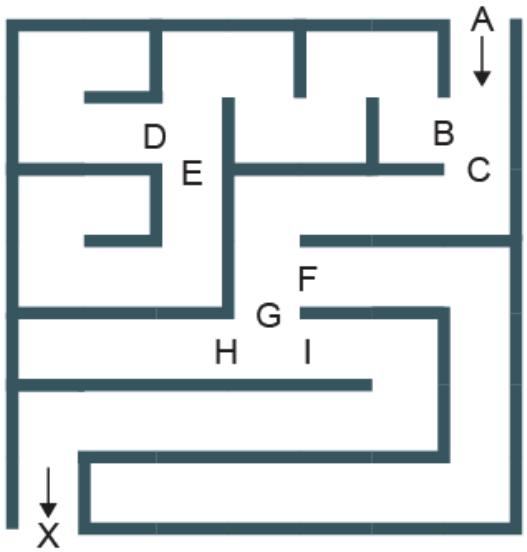
    A → B → C is 5 + 10 = 15

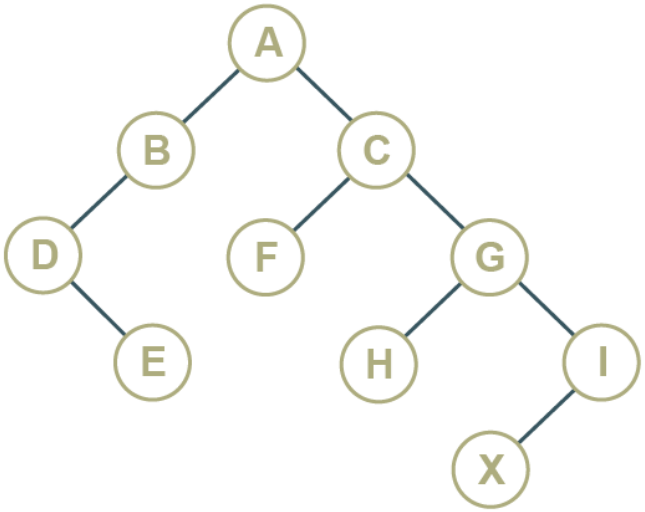
    A → B → D is 5 + 18 = 23
  - The shortest path so far is A → B → C = 15
  - Backtrack to C. A → B → C → D = 15 + 6 = **21**



Solving mazes

- A maze can be represented as a graph
- A depth-first traversal is used to explore a path until a dead-end is reached, and then it backtracks to the previous node, and tries the next path





Worksheet 6

- Try the first question in **Task 1** of Worksheet 6
  - If you have time, try the second one too

Can it be solved?

- Sometimes it is not possible to find a solution to a problem in a reasonable time
- Cracking a strong password may be impossible using an exhaustive search

Cracking a password

- What other methods, apart from an exhaustive search taking 34 thousand years, could you try to discover someone's password or crack a code?

Educated guess?

- To crack a password, you could try guesswork based on the person's family, names, birthdates, favourite team, etc.
- You could try comparing it with words in a dictionary
- To crack a code, if you had a large chunk of encrypted text, you could use a frequency analysis of letters – for example 'e' is the commonest letter in English, followed by 't'

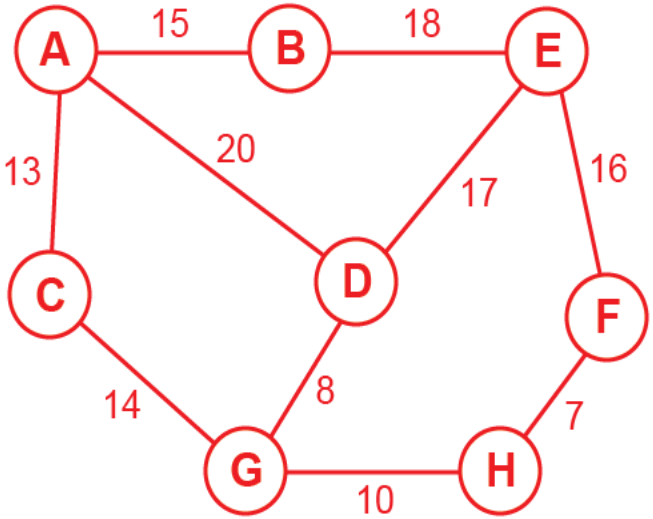
chess.

Heuristic methods

- Heuristic methods include ‘rules of thumb’, educated guesses, intuitive judgements or common sense
- Teachers, students, psychologists, chess players, use heuristics to solve problems
  - How often do you use a heuristic approach to a problem? Think of examples

An intractable problem

- The Travelling Salesman Problem (TSP) is a problem which cannot be completely solved
- The problem is to find the shortest route which visits every city and returns to the start



Heuristic methods

- A heuristic method is used to rapidly find a solution that is ‘good enough’, even though it might not be the optimal solution
- Many algorithms have been invented to provide a heuristic solution to the Travelling Salesman Problem (TSP)
  - The best heuristic solutions are within 2-3% of the optimal solution for up to 85,000 cities or nodes

Applications of heuristic methods

- The heuristic approach is used in many applications such as
  - routing messages across the Internet
  - building circuit boards
  - transportation
  - virus checking
  - DNA analysis
  - artificial intelligence

Data mining

- Data mining is the process of collecting and then analysing huge amounts of data
- Many organisations such as the National Health Service, police, Google, Amazon and supermarkets collect billions of bytes of data about people
- They can then analyse or ‘mine’ the data to find connections and associations
- A range of modelling techniques are used to help to identify patterns in the data

Applications of data mining

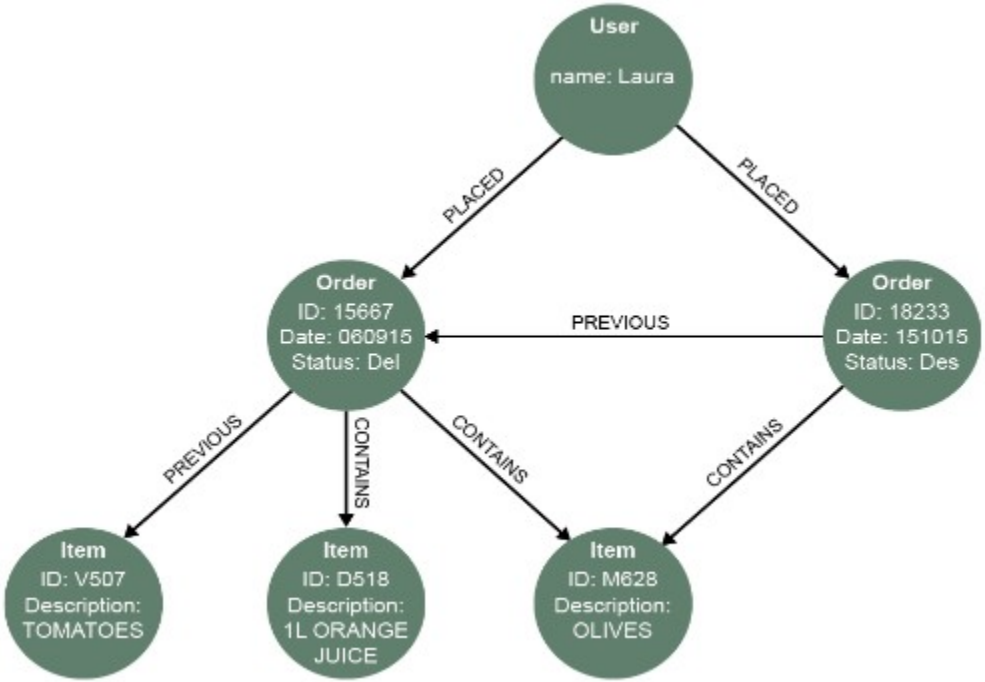
- Increasing response rates to marketing campaigns by being able to target them more accurately to the needs of each customer
- Anticipating resource demands
- Detecting fraud and cybersecurity issues
- Finding connections between seemingly unconnected events

Big data

- The term big data is often used in connection with data mining
- It implies that huge amounts of data are collected and stored
- It is defined by three major features, known as the 3Vs: Volume, Variety and Velocity
- Parallel computing, in which algorithm tasks are executed concurrently on a cluster of machines or supercomputers, is fundamental to managing big data tasks

Mining supermarket data

- A supermarket can customise their advertising and special offers, depending on your purchase history or that of other customers who

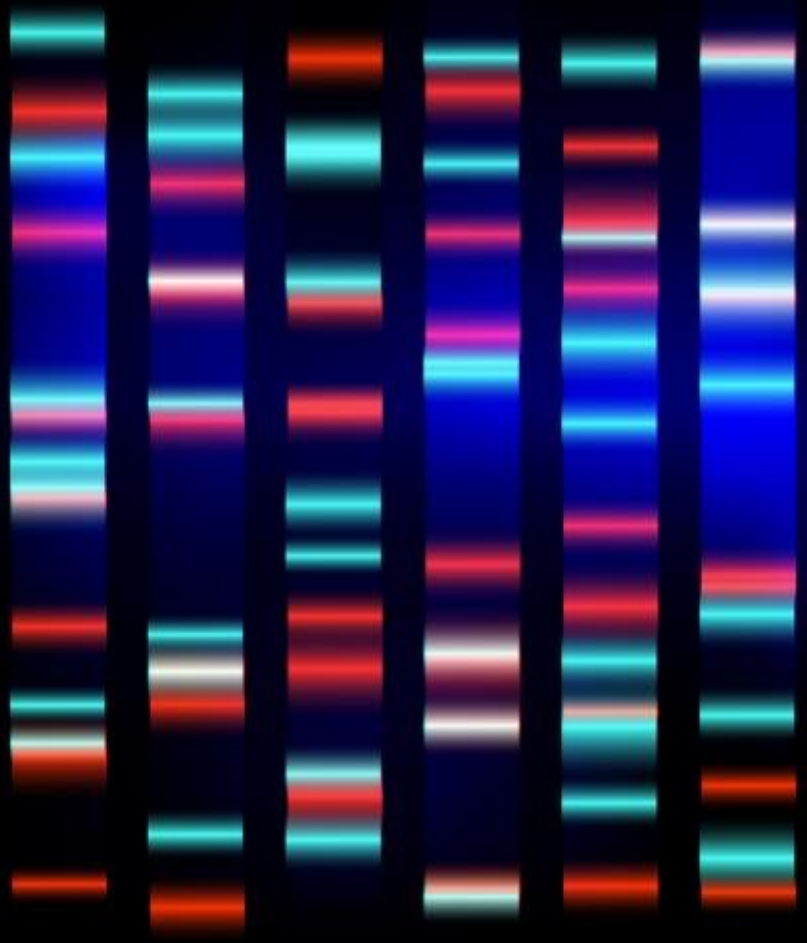


Amazon data mining

- Amazon make recommendations based on what other readers who have read the same books as you have also read

Medical applications of data mining

- Data mining and big data techniques are widely used in the public health field, discovering new patterns among population groups using social media data



Performance modelling

- How efficient is your algorithm?
- One measure of efficiency is the Big-O notation (covered in Section 12)
- This measures the suitability of algorithms in terms of execution time and space
- You can time different algorithms using built-in timer functions
  - The performance of some algorithms deteriorates so rapidly as the problem size increases that they are not usable

Performance of sorting algorithms

- Sorting algorithms vary enormously in their efficiency
- The bubble sort is the slowest of the well-known algorithms
- The quicksort is very fast for a large number of items

Worksheet 2

- Now try **Task 2** in the worksheet

Pipelining

- Pipelining is an implementation technique where multiple instructions are overlapped in execution
- Instructions enter the ‘pipeline’ at one end, and at each stage part of the instruction is completed and moves to the next stage while another instruction enters the pipeline – rather like an assembly line

Pipelining in PCs

- Pipelining is now commonplace in PCs
- Intel chips can execute many instructions simultaneously to achieve high processing speeds of 3GHz and more

Plenary

- You should be able to explain and give examples of
  - visualisation
  - backtracking
  - heuristics
  - data mining
  - performance modelling
  - pipelining

  and say how they are used in solving problems

Copyright

Licence agreement